DOCUMENT RESUME

ED 298 559 CS 506 331

AUTHOR Kennamer, J. David
TITLE Often, Sometimes, Rarely or Never: Issues in the Measurement of Communication Behaviors.
PUB DATE Jul 88
NOTE 42p.; Paper presented at the Annual Meeting of the Association for Education in Journalism and Mass Communication (71st, Portland, OR, July 2-5, 1988).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Communication Research; *Data Collection; *Data Interpretation; Higher Education; Measurement Techniques; *Public Opinion; *Rating Scales; *Statistical Surveys
IDENTIFIERS *Communication Behavior; *Quantifiers; Verbal Labeling

ABSTRACT

"Vague quantifiers" used on quasi-quantitative rating scales (such as "often, sometimes, rarely, or never") are discussed in this paper in terms of the problems they create both for theory development and for issues of analysis. The purpose of the paper was to investigate some of these issues as they relate to the measurement of communication behavior, specifically, newspaper and television news use and political discussion. The paper first examines two public opinion surveys on political campaign discussion. Both surveys were conducted in Richmond, Virginia--the first resulted in 217 completed interviews, and the second in 324 interviews. The approach to the analysis was to use the choice of verbal labels as the basis for forming groups. The paper then concludes from the data that many measures of critical variables are (1) weak; (2) unclear in meaning both to the researchers and the research subjects; (3) unclear in their reliability and validity; (4) ambiguous in terms of their place on the nominal to ratio measurement continuum; and (5) prone to huge amounts of error. Ten tables of data and four figures are included, and 13 references are appended. (MS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J David Kennamer
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED298559

OFTEN, SOMETIMES, RARELY OR NEVER:

ISSUES IN THE MEASUREMENT OF COMMUNICATION BEHAVIORS

BY

J. David Kennamer

School of Mass Communications
Virginia Commonwealth University
Richmond, Virginia
23284

Submitted to the Communication Theory and Methodology Division, AEJMC

CS506331

BEST COPY AVAILABLE

OFTEN, SOMETIMES, RARELY OR NEVER:

ISSUES IN THE MEASUREMENT OF COMMUNICATION BEHAVIORS

J. David Kennamer
School of Mass Communications
Virginia Commonwealth University
Richmond, Virginia 23284
(804) 367-1260

Respondents in two surveys were asked to provide numerical estimations of the "vague quantifiers" often used to measure frequency and amount of social science concepts. Among those examined here were the use of "often, sometimes, rarely or never" to describe political discussion and news media use.

The means of these estimations indicate that the verbal labels do not closely approximate an interval scale. In addition, large standard deviations indicate serious overlap in their meaning to respondents. The number of scale points and the presence of a ceiling also affect response distributions.

Theory and Methodology Division
Session Title: "Getting Involved with Communications"

OFTEN, SOMETIMES, RARELY OR NEVER: ISSUES IN THE MEASUREMENT OF COMMUNICATION BEHAVIORS

> "Whether the water in a pot feels cold, cool, warm or hot depends both on how long the pot has been on the fire and how long the resting hand has been in from the cold" (Shepard, 1978).

Survey respondents are often asked to make judgments of magnitude or frequency of a behavior, belief, expectation, perception, emotion, or attitude on quasi-quantitative rating scales such as "often, sometimes, rarely, never," "strongly approve, somewhat approve ...," "too little, about right, too much."

The meaning of such "vague quantifiers" (Bradburn and Miles, 1979) is, of course, problematic, both for theory development and for issues of analysis. Are they ordinal, interval, or even ratio scales? And what, if anything, do they really mean to respondents? And what if the meaning of the concept being measured with these vague quantifiers is itself only vaguely understood by the respondent? Thus, the use of such terms involves issues of both reliability and validity, it seems.

The purpose of this paper is to investigate some of these issues as they relate to the measurement of communication behavior, specifically, newspaper and television news use and political discussion. Nearly every convention session includes the complaint by a reviewer, discussant or audience member about the weakness of our measures of media use and other forms of communication behavior, especially as they relate to politics and elections. This paper will attempt to shed some light on

the nature and magnitude of the problem, and will also make some suggestions that might lead to improvement.

OFTEN, SOMETIMES, RARELY,...(uh), MAYBE?

In one of the few studies that have addressed this issue, Bradburn and Miles (1979) looked at an exceedingly slippery "feeling state," how often in the past few weeks the respondents reported feeling particularly excited or bored about something, with the response categories of "never, not too often, pretty often, or very often." They then asked respondents to provide a precise numerical estimation of the frequency of this feeling, which the researchers then transformed into times per month. Thus, their research seems to be an extreme case of using vague quantifiers to measure feelings that are not only difficult, but probably impossible, to quantify. In addition, no firm time frame is given, no topic or issue area is referred to (nothing to be excited or bored about), and it is hard to imagine a social norm providing much guidance.

Therefore, it should come as no surprise that Bradburn and Miles found considerable "fuzziness" in the relationship between the numerical estimates and the verbal labels. In addition, it seems that the precision provided by such numbers is really false. This falseness may derive from two sources: The labels are themselves unclear in what they imply for measurement (is "often" five times a week or five

times a year?). And the variable being rated or counted may not be amenable to "counting."

The meanings of the adverbs and adjectives used to rate or count social science concepts have been studied sporadically and not very systematically, it seems, for several decades. Mosier (1941) proposed that the meaning of a word is made up of three components: a meaning that is constant across individuals and situations, i.e., its inherent meaning, variations in meaning across individuals, and variation due to the context in which the word is found. Despite these complications, he found that numeric estimations of the favorableness-unfavorableness implied by a modifier tended to assume a normal distribution.

Cliff (1959) obtained numeric estimations of a large number of adverbs common in general use and in social science questions. His purpose was to see how much the adverbs changed the meaning of adjectives, and to come up with approximate multipliers for these adverbs. His results indicated for example, that "in a very real sense, 'extremely good' is about one-and-one-half times as good as 'good'." (1959, p. 42). In fact, assuming a value of 1.00 for an unmodified adjective, he was able to develop from respondent reactions to these variables consistent multiplicative values for a set of adverbs ranging from slightly (.55), somewhat (.68), rather (.85), pretty (.91) quite (1.05), decidedly (1.20), unusually (1.30), very (1.30), and extremely (1.5).

Similarly, Bass (1968) asked student subjects to assign a reference number to "sometimes" and then indicate what number would best fit 28 other adverbs, given the value they had assigned to "sometimes." He discovered that the terms "always," "very often," "fairly often," "sometimes," "seldom" and "never" bore an approximate relation to each other of 5:4:3:2:1:0. He then used these in his questionnaire items.

A 1974 replication by Bass, Cascio, and O'Conner found that manipulation of the importance of the topic being asked about did not affect the estimations of frequencies described by the modifiers. Others have argued, however, that the context in which such modifiers are embedded may have a major effect on their meaning to respondents (Parducci, 1968). Chase (1969) developed separate scales, one with frequency modifiers weighted toward low frequencies, and another with predominately high frequency terms. He found that respondents tended to respond to them in the same fashion, leading him to conclude that, "Respondents get a good deal of meaning from scale adjectives because of the adjective's relative position in a group of response categories, rather than in terms of a 'standard' definition of a given word out of context of other category labels" (p.1043).

The findings of Pepper and Prytulak (1974) seem to confirm the importance of context effects. When a topic was perceived as being of low frequency (air crashes,

earthquakes) numerical definitions of terms describing those activities were considerably lower than when topics were of high frequency (missing breakfast, gunfire in Westerns). Aside from their objective frequency, these topics also seem to differ in the importance or seriousness of their outcomes.

These studies seem to indicate that these commonly used quantifiers are problematic. While some researchers have found that some of them seem to have core meanings, these meanings seem to be rather easily diverted by the context of presentation or topic being rated.

The other factor discussed by Bradburn and Miles is simply the variable's inherent "measurableness." They write, "While it is relatively easy for respondents to report how many times in the past week they have been to a movie, or how many hours they watched television yesterday, they seem to have a great deal of difficulty in putting precise numbers on subjective states and on events of relatively low salience" (1979, p. 100).

Their comments on subjective states seem to be self evident, and perhaps more important to areas of psychology and social psychology than mass communications research. However, they make no further reference to these "low salience" events, and this seems to be a highly relevant point since many, perhaps most, communications behaviors could be classified in this manner. It is a premise of this paper that many behaviors are not particularly "salient" for a variety of reasons. They may

be in some sense "scripted" and therefore not thought much about when they are enacted. They may not involve clear decisions to act but are simply reactions to the initiatives of other people. Perhaps they have unclear boundaries and are not easily distinguishable from other activities, or occur as part of other activities and are not "coded" in memory separately.

For example, at a recent party to celebrate Superbowl XXII, conversation ranged over a variety of topics, only some of which had much to do with the football game. A comparison was made between the flared tempers on the playing field to the Dan Rather-George Bush fracas that had broken out the previous week on CBS Evening News. A ten-minute conversation ensued concerning the presidential election, the upcoming "Super Tuesday" Southern primary and the merits of various candidates. Someone then complimented the host on the clam dip and the conversation moved on from there. This was clearly a "political discussion" viewed from outside, but to what extent was it coded as such by the participants and remembered? This may be what is meant by an event of low salience, and it might be predicted that respondents would have some difficulty in recalling such events with much reliability.

Other communications behaviors may be more easily measured. For example, watching television news or reading a newspaper clearly involves the manipulation of physical objects and the assumption of a particular physical

position. They may both relate to long-established habits and arrangements such as the delivery of a newspaper in the morning or stopping at the 7-11 for coffee and a New York Times on the way to work. Both, perhaps television less so than newspaper reading, involve a conscious choice. Individuals may be much more able to remember such activities accurately than a more abstract concepts such as discussion of politics.

A generalization to be derived from this discussion, therefore, is that behaviors probably can be rated along a continuum from the relatively concrete and discrete to those that are very nearly as abstract and unobservable as a feeling or attitude. One would expect, therefore, that efforts to quantify the former would be relatively easy, that people would be able to reliably count these events and report them. In the latter case, however, efforts to count such behaviors would probably provide no more precision than what can be obtained with vague quantifiers.

In the analyses that follow, "vague quantifiers" are used to measure the frequency of several communications behaviors. Then attempts are made to assess the degree to which respondents are able to express them in numerical terms.

DATA AND RESEARCH DESIGN

The investigation of this issue follows the same method used by Bradburn and Miles. That is, respondents to a

survey were asked to estimate the extent of a behavior using a scale made up of vague quantifiers. In two instances they were then asked to apply a precise numerical estimation to their selection of a vague quantifier. In other instances separate and independent questions were used to try to provide a numerical estimation of the actual frequency of the behavior. The analysis largely consists of comparisons of means and standard deviations of the numerical estimations across the sets of vague quantifiers.

The data come from two surveys, both conducted by students in an undergraduate class in public opinion in the Richmond, Virginia metropolitan area. The most recent, and the one most heavily used in this analysis, was conducted in October, 1987 and resulted in 217 completed interviews. The second set of data consists of 324 interviews from a presidential election survey conducted in October 1984. Selection of respondents in both samples was based on random selection from the Richmond area telephone directory. In both cases a random-digit, add-on procedure was used to modify numbers selected from the telephone book. Both directory numbers and the randomly modified numbers were used, thus ensuring some possibility of obtaining interviews with individuals who were not listed in the directory.

Political discussion was measured in the same fashion in both the 1984 and 1987 surveys with two questions. The basic question asked "How often do you talk about politics

or public issues with other people, often, sometimes, rarely or never?" This was then followed, for those who gave a response other than "never," by the following question: "When you say you talk about these issues (response to previous question), about how many times a week would you say that is?"

In addition, the 1984 survey contained four other questions that can be used to assess the amount of specific campaign discussion independently of these two questions. Respondents were asked if they remembered "taking part in conversations in which the (presidential) debates were discussed." A similar question asked about participation in conversations about the nominating conventions. In addition, respondents were asked if they had "tried to convince someone to vote for the candidate you favor" and if they had been the target of such persuasive attempts. All four questions were answered in a simple yes or no format. The "yes" responses were summed to create a campaign discussion index with a range of 0 to 4.

The 1987 survey also contained a number of efforts to measure exposure and attention to political news, both in newspapers and on television. For newspaper use, the basic "vague quantifier" question is "How often do you read stories about world or national events... often, sometimes, rarely or never." A number of questions attempted to assess newspaper use, but the most relevant here seem to be: "How many days in the past week did you read the front page of a

newspaper?" This was followed, for those who said something other than "none," with: "Think now about how much of the front page you read on an average day. Where would you place yourself on a scale of 1 to 10, where 1 means you scan the headlines, to 10 which means you read all of every story?"

Clearly, this is not equivalent to the two discussion questions when the vague quantifiers and the numerical estimations refer to the exact same behavior. In these two cases of newspaper use, however, numerical estimations are attempted of behaviors that seem to be necessary if reading about world or national events is to be accomplished: frequency of reading the front page of a newspaper, and amount of the front page read.

Television use and exposure are more difficult issues to assess. In this case, the basic question is, "When you watch the news, would you say you give it your full attention, some attention, or not very much attention?" This was followed by: "Many people do other things while the TV news is on. When you have the news on, are you usually doing something else, sometimes doing something else, or rarely doing something else?" This is, of course, trying to validate one set of vague quantifiers with another, although in the second instance, the question refers to specific concrete behaviors, events and habits.

This assumption seems to be a valid one given the ease with which respondents answered a following questions asking

what those other activities were. Respondents reported fixing dinner, cleaning house, changing diapers, reading, sewing or knitting, and talking to other people.

ANALYSIS AND RESULTS

The approach to the analysis is to use the choice of verbal labels--often, sometimes, rarely, never--as the basis for forming groups. Oneway ANOVA is used to test for mean differences in numerical estimations of the behaviors across the groups defined by the vague quantifier ratings of those behaviors. In addition to tests of mean differences, tests for homogeneitiy of variance across the groups are also provided. Other analyses try to evaluate the distance between the group means in terms of percent change from one mean to the next. In addition, the distributions described by the means and standard deviations are displayed in graphs. Tables 1-10 and Figures 1-4 show the results.

Table 1 shows the mean times per week discussing politics by the Often, Sometimes or Rarely responses to a previous question[1] for the 1984 and 1987 samples. The

---

TABLE 1 ABOUT HERE

---

patterns are the same for both years. The means, both in magnitude and degree of change, are remarkably similar. As would be expected, those who said they discuss politics

often report a much higher numerical estimation of this behavior than those who discuss sometimes. These people report greater frequency than those who report discussing politics rarely.

In addition, there seems to be a fairly regular decline in the size of the standard deviation in frequency of discussion from the Often category to the Rarely category. This is clearest in the 1984 data. The statistical tests show that for both years the mean differences are statistically significant ($p < .05$), and that the standard deviations are also significantly different from each other.

Figure 1 shows the means on a continuum with the range defined by one standard deviation plotted to either side of each mean. The large degree of overlap in these ranges clearly demonstrates the "fuzziness" of these vague quantifiers. Note that there is some overlap in both cases of the Often and the Rarely groups. Also it can be seen that the verbal labels are not equidistant on the scale. In both cases "sometimes" is closer in meaning to "rarely" than to "often." In fact they seem to describe a roughly 1:2:4 relationship.

FIGURE 1 ABOUT HERE

Table 2 shows similar analyses for the campaign discussion index for the 1984 sample. Again, the score on

this index increases monotonically as the magnitude of the verbal labels increases from never to often.[2] Whereas in the other discussion analyses, the standard deviation decreased with labels of lesser magnitude, here it increases. The standard deviation is 0.79 for the Often category and increases monotonically to 1.29 for the Never category. The differences among the means and standard deviations are statistically significant.

TABLE 2 ABOUT HERE

Figure 2 graphs the ranges that are defined by the means plus or minus one standard deviation. Again, an extreme degree of overlap is seen. On this four-point scale, there is even considerable overlap of the Often and Never groups in the actual numeric frequency of the behavior. And again, the verbal labels do not define equal distances on the scale. The distance between "rarely" and "sometimes" is larger than the distances between "sometimes" and "often" and between "sometimes" and "never."

FIGURE 2 ABOUT HERE

Table 3 shows results for the days in the past week the respondents reported reading the front page of a newspaper by whether they "often, sometimes, rarely or never" read about national and world events.[3] There are large

differences in the means here, with the Often group reporting 6.02 days, the Sometimes group 4.66 days and the Rarely/Never group 2.81 days. The standard deviations show the same pattern as was seen with the discussion index (Table 2); they increase as the the magnitude of the verbal

---

TABLE 3 ABOUT HERE

---

labels decreases. Both mean and standard deviation differences are statistically significant.

Table 4 shows the mean differences in amount of front page read (1 to 10 scale) by the same verbal labels for frequency of reading about world and national events. The same pattern in the means is found here as in the previous analysis. There is also a hint of the same pattern in the standard deviations but it is much less clear.

Figure 3 graphs these two analyses and demonstrates again the degree of ambiguity in their quantitative meanings.

TABLE 4, FIGURE 3 ABOUT HERE

Finally, Table 5 shows the mean frequency of "doing other things" by attention paid to television news as measured by the "full," "some," and "not much" attention labels. Again, the numbers are the mean of the codes assigned to the "usually, sometimes, or rarely" verbal

labels assigned to the "doing other things" measure. Clearly, people who say they give full attention report doing fewer other activities than people who report paying "some" attention, and these people engage in fewer activities than those who report paying "not much" attention. The standard deviation declines in magnitude as the magnitude of these verbal labels declines, as was the case with the first two discussion measures (Table 1).

Figure 4, showing the means with the ranges defined by adding and subtracting one standard deviation from each, again demonstrates considerable overlap.

TABLE 5, FIGURE 4 ABOUT HERE

Comparisons among these various tables are difficult since they differ in the nature of the variable, possible range of responses and in number of response categories. One way to approach this is to compare percent change in means as one moves from the lowest magnitude response to the highest. Tables 6 through 10 shows these calculations, including a mean percent increase across all the response categories.

Table 6 shows that for the discussion questions, the mean percent increase is 117 percent for the 1984 sample, and 104 percent for the 1987 sample. The remainder of these differences are much smaller. For the discussion index, the change is, on average, 36 percent (Table 7). For the two

measures of front page reading, they are 48 percent (Table 8) and 27 percent (Table 9). Table 10 shows that the average percent change in the frequency of "doing things" while watching television news is 24 percent.

TABLES 6 - 10 ABOUT HERE

DISCUSSION

These data seem to confirm what many of us have always suspected, that many of our measures of critical variables are weak, unclear in meaning both to us and our research subjects, unclear in their reliability and validity, ambiguous in terms of their place on the nominal to ratio measurement continuum, and prone to huge amounts of error.

Much of the research cited earlier deals with efforts to measure psychological states--attitudes and perceptions--and one should perhaps expect that measuring such internal and unobservable constructs would be difficult. This study, however, looks at more or less concrete behaviors, and seems to find the same problems of meaning and reliability. This is particularly problematic because our major independent variables tend to be communication <u>behaviors</u>. In theory, it seems that a behavior should be easier to measure reliably than an internal "feeling state" because they are more easily observable and countable, either by those who do the

behaving or by outside observors of behavior. However, this study seems to indicate that the measurement difficulties plaguing "feeling states" also affect the measurement of the frequency and intensity of behaviors.

This is clearly indicated here by the large standard deviations associated with all the means reported in these ANOVAs. In several instances there seems to be little difference to repondents on average in their numerical estimations of the meaning of "sometimes" and "often. There is even significant overlap in the estimations of the meaning of "rarely" and "often."

We also see that these "vague quantifiers" are not equally spaced in the numeric estimates. "Sometimes" seems to be closer in meaning to "rarely" than to "often." In addition we see significant heteroscedasticity in the variances of the various groups with some indication of systematic differences depending upon the nature of the behavior being estimated.

It was noted earlier that behaviors may differ in their concreteness and "countability." It is very hard to evaluate this notion with these outcomes. The basic discussion questions (Table 1 and Figure 1) show the largest percent increases in means, which would seem to indicate that they are more discriminable, which is the opposite of what might be expected if one assumes these are relatively abstract and hard to remember behaviors. However, the scale upon which respondents estimated these frequencies was

completely open; they could answer anything. Estimates of television and newspaper use were constrained at the upper end by a definite scale value (1-10, 0-7), as was the index of discussion (0-4). This in itself may account for the large differences in these discussion frequency means.

This is a flaw in the research design, obviously, but it also demonstrates how the operationalization of variables can make such a difference in the responses that are obtained and in the potential outcomes of analysis. This is clearly shown in the two operationalizations of discussion. One is a simple numerical estimation of the frequency of the behavior. The other is an index made up of answers to independent questions concerning a group of concrete behaviors. The outcome of the latter is much closer to the patterns shown by the television and newspaper analyses. The means show similar levels of percent change across increases in the magnitude of the verbal labels, and the pattern in the standard deviations is the same: the magnitude of the standard deviations decreases as the magnitude of the means increases, consistently across several measures. So in one case, discussion is operationalized as an abstraction and in another as a sum of concrete behaviors.

In the case of these more concretely conceptualized or measured behaviors, this pattern in the standard deviations could easily occur because the people in the "often" category reported on firmly established and very clear

habits, because of a ceiling effect created because they are all at the top of the scale, and because the people in the "rarely" and/or "never" category responded according to social norms or somewhat randomly because of their lack of connection with the entire issue. These may be particularly "non-salient" behaviors for some people.

Why, then, is the pattern of standard deviations so different on the more abstract discussion measures? For one thing, there is no ceiling provided by a scale, so respondents in the "Often" group could respond in any fashion they wanted. Also, there are no clearly defined habits here, such as a newspaper subscription, or eating dinner during the evening news, which would tend to standardize responses across large numbers of people. People in the "rarely" category might, given the undefined and abstract nature of the behavior, simply opt out of the question with a response of one or two times per week. In fact in the 1984 sample, 75 percent of the "rarely" group responded with once per week. In 1987, this was 80 percent.

IMPLICATIONS FOR RESEARCH

While the actual processes lying behind these patterns of means and standard deviations remain largely speculative, the existence of the measurement problems has been made quite clear. Do the results of this study provide any

guidance for research? It may not be enough simply to document the nature of the problem.

The research cited earlier in this paper probably deserves wider dissemination and understanding. Efforts have been made by various researchers to find out what these vague quantifiers mean to people, and then to select ones that bear the proper relationship to each other, that form roughly interval scales, for use as response categories. While some researchers seem to think the meanings of these adverbs and adjectives are stable enough that standard lists of these terms can be presented across contexts, others are concerned that the context of their presentation could easily change their meaning. If that is the case, then researchers could include in their pretests investigations of the meaning of various response categories within the context of the research being contemplated.

Bradburn and Miles, on the other hand, discovered that the modifiers they used did not form an interval scale. In other words, assigning standard codes of say, 1, 2, 3, and 4 to the vague quantifier categories was inappropriate because these the meanings of the terms did relate to each other in this clean linear fashion. Instead they attempted to weight the codes in relation to the numerical estimations actually given to the words by respondents. Instituting this procedure would require, however, that researchers attempt, within the research, to determine the meaning of their response categories to respondents.

It is also possible that the volumnious research tradition in psychophysics could be tapped in this regard. Researchers in this area have determined that individuals are able to map with considerable discrimination the magnitude of one physical stimulus onto another another physical modality, and then, the magnitude of a social or psychological "stimulus" in terms of this physical dimension. Cross (1982) describes the basic approach and basic results of these experiments:

> ...if you present individuals with a sequence of stimuli that vary in magnitude along one perceptual or cognitive dimension and request that they "track" or respond to these changes by producing analogous changes in a different dimension, the task can usually be performed even when neither variable is given directly in numerical value. For example a person can indicate his/her opinion of the desirability of various occupations by adjusting the intensity of a light, or the level of a noise burst, so that the levels of brightness and loudness produced are proportional to the desirability of each occupation. (p. 75).

This would result, therefore, in a ratio level scale for this very basic sociological variable, occupational prestige. Lodge (1981) notes that "probably the most strongly supported law of human judgments in psychology today" is that "equal stimulus ratios produce equal subjective ratios" (p. 12-13).

While it is in most cases impractical for survey researchers to validate measures in this fashion, this research seems to indicate that such abstract conctructs <u>can</u> be measured with reasonable reliability on ratio or interval scales. Borgatta and Bohrnstedt (1980) point out that most

variables of interest to social scientists are "latent unobserved constructs" which "are conceptualized to be continuous at the latent level, even though they are usually manifestly measured as discrete variables....If the constructs are continuous, they must also be interval" (p. 152). They note that social science researchers usually do a poor job of assessing or matching these underlying distributions with measures of the sort investigated in this paper. Nonetheless, despite the appearance of these measures, they insist they are still interval because the latent variable is interval. "At the manifest, observed level, our measures are likely to be imperfect interval-level scales. Social measurement is crude...." (p. 153). The degree of imperfection is simply measurement error, and does not necessarily degrade the measure from interval to ordinal, these authors argue.

This paper has attempted to describe just how imperfect some of our measures actually are. It is important to remember that our measures are usually several steps removed from the construct we are measuring. We ask respondents to rate abstract constructs in terms of vague verbal labels. Then we attach numerical values to the verbal labels, and then combine the numerical values in various ways and call the results something else. Marks (1982, p. 44) points out that this is a metaphorical process, "In a sense, comparing the validity, reliability, or whatever other properties of different scaling methods is much like saying: 'What are the good metaphors'?"

NOTES

1. Respondents who answered "never" to the first "vague quantifier" question were not asked to estimate the times per week question.

2. Since the questions forming the index were completely independent of the "vague quantifier" question, the mean for the "never" group is included.

3. Because only 6 people said they never read about national or world events, the "never" and "rarely" groups were combined for these analyses.

REFERENCES

Bass, Bernard M. (1968) "How to succeed in business according to business students and managers." Journal of Applied Psychology, 52:3:254-262.

Bass, Bernard M., Wayne F. Cascio, and Edward J. O'Connor (1974) "Magnitude estimations of expressions of frequency and amount." Journal of Applied Psychology, 59:3: 313-320.

Borgatta, Edgar F., and George W. Bohrnstedt (1980) "Level of measurement, Once over again." Sociological Methods and Research, 9:2:147-160.

Bradburn, Norman M., and Carrie Miles (1979) "Vague Quantifiers." Public Opinion Quarterly, 43:1:92-101.

Chase, Clinton, I. (1969) "Often is where you find it." American Psychologist, 24:11:1043.

Cliff, Norman (1959) "Adverbs as multipliers." Psychological Review, 66:1:27-44.

Cross, David V. (1982) "On judgments of magnitude." In B. Wegener (ed.) Social Attitudes and Psychophysical Measurement. Hillsdale, NJ: Erlbaum.

Lodge, Milton (1981) Magnitude Scaling: Quantitative Measurement of Opinions. Beverly Hills: Sage.

Marks, Lawrence E. (1982) "Psychophysical Measurement: Procedures, Tasks, Scales." In B. Wegener (ed.) Social Attitudes and Psychophysical Measurement. Hillsdale, NJ: Erlbaum.

Mosier, Charles I. (1941) "A psychometric study of meaning." Journal of Social Psychology, 13:123-140.

Parducci, Allen (1982) "Category ratings: Still more contextual effects!" In B. Wegener (ed.) Social Attitudes and Psychophysical Measurement. Hillsdale, NJ: Erlbaum.

Pepper, Susan and Lubomir S. Prytulak (1974) "Sometimes frequently means seldom: Context effects in the interpretation of quantitative expressions." Journal of Research in Personality, 8:95-101.

Shepard, Roger N. (1978) "On the status of 'direct' psychophysical measurement." in C.W. Savage (ed.) Perception and Cognition Issues in the Foundations of Psychology. Minneapolis: University of Minnesota Press.

Table 1

Days Per Week Discussing Politics by
"often, sometimes, or rarely" labels
Means, Standard Deviations and Ranges for
1984, 1987 Samples

	Mean		SD		Count		Range	
Vague Quantifier	1984	1987	1984	1987	1984	1987	1984	1987
Often	5.80	5.84	3.70	2.86	80	50	1-25	2-14
Sometimes	2.61	2.92	1.69	3.07	125	79	1-14	1-20
Rarely	1.40	1.55	0.90	1.93	65	56	1-5	0-9
Total	3.26	3.30	2.91	3.16	270	185	1-25	0-20

1984: $F(2,267) = 71.33$; $p = .0000$
Homogeneity of Variance
Cochran's C = 0.79; p =.000
Bartlett-Box F = 67.16; p =.000

1987: $F(2,182) = 34.31$; $p = .0000$
Homogeneity of Varience
Cochran's C = .44; p =.049
Barlett-Box F = 6.58; p =.001

Table 2

Campaign Discussion Index by Discussing Politics
"often, sometimes, rarely or never"
Means, Standard Deviations, and Ranges for 1984 Sample

Vague Quantifier	Mean	SD	Count	Range
often	2.71	0.79	82	1-4
sometimes	2.24	1.12	139	0-4
rarely	1.44	1.19	77	0-4
never	1.11	1.29	19	0-4
Total	2.10	1.18	317	0-4

F(3,313) = 24.86; p = 0.0000
Homogeneity of Variance
Cochran's C = .34; p = 0.035
Bartlett-Box F = 5.103; p = .002

Table 3

Days in Past Week Read Front Page
by Reading about National and World Events
"often, sometimes, rarely or never"
Means, Standard Deviations and Ranges for 1987 Sample

Vague Quantifiers	Mean	SD	Count	Range
Often	6.02	1.81	99	1-7
Sometimes	4.66	2.33	47	0-7
Rarely/never*	2.81	2.77	37	0-7
Total	5.02	2.50	183	0-7

F (2,180) = 30.32; p = .0000
Homogenity of Variance
Cochran's C = .4672; p = .014
Bartlett-Box F = 5.552; p = .004

*The "never" and "rarely" categories are combined because the "never" category by itself contains only 6 respondents.

Table 4

Amount of Front Page Read by Reading about National and World Events "often, sometimes, rarely or never"
Means, Standard Deviations, and Ranges for 1987 Sample

Vague Quantifiers	Mean	SD	Count	Range
Often	6.04	2.20	98	1-10
Sometimes	4.59	2.04	46	1-10
Rarely/Never	3.82	2.60	28	1-10
Total	5.29	2.39	172	1-10

F (2,169) = 13.93; p = .0000
Homogeneity of Variance
Cochran's C = .4283; p = .107
Bartlett-Box F = 1.040; p = .354

Table 5

Frequency of "doing other things" by
"full, some or not much" Attention Paid to Television News
Means, Standard Deviations and Ranges for 1987 Sample

Vague Quantifiers	Mean	SD	Count	Range
Full	1.64	0.72	85	1-3
Some	2.36	0.64	90	1-3
Not much	2.83	0.41	6	2-3
Total	2.03	0.77	181	1-3

F(2,178) = 13.33; p = .0000
Homogeneity of Variance
Cochran's C = .4736; p = ..011
Bartlett-Box F = 1.484; p = .227

Table 6

Percent Change in Frequency of Discussion
by "often, sometimes, rarely" levels

	1984			1987		
	Mean	Diff	Percent Change	Mean	Diff	Percent Change
Often	5.80			3.70		
		3.19	147%		2.01	119%
Sometimes	2.61			1.69		
		1.21	86%		0.79	88%
Rarely	1.40			0.90		
Total (mean)	3.26	2.22	117%	3.30	1.05	104%

Table 7

Campaign Discussion Index by
"often, sometimes, rarely, never" labels
1987 Sample

	Mean	Difference	Percent Change
Often	2.71		
		0.47	21%
Sometimes	2.24		
		0.80	56%
Rarely	1.44		
		0.33	30%
Never	1.11		
Mean		0.53	36%

Table 8

Percent Change in Frequency of Front Page Reading
by "often, sometimes, rarely or never" labels
1987 Sample

	Mean	Difference	Percent Change
Often	6.02		
		1.36	29%
Sometimes	4.66		
		1.85	66%
Rarely/Never	2.81		
Mean	5.02	1.61	48%

Table 9

Percent Change in Front Page Reading by
"often, sometimes, rarely, or never" labels
1987 Sample

	Mean	Difference	Percent Change
Often	6.04		
		1.45	32%
Sometimes	4.59		
		0.77	21%
Rarely/Never	3.82		
Total	5.29	1.11	27%

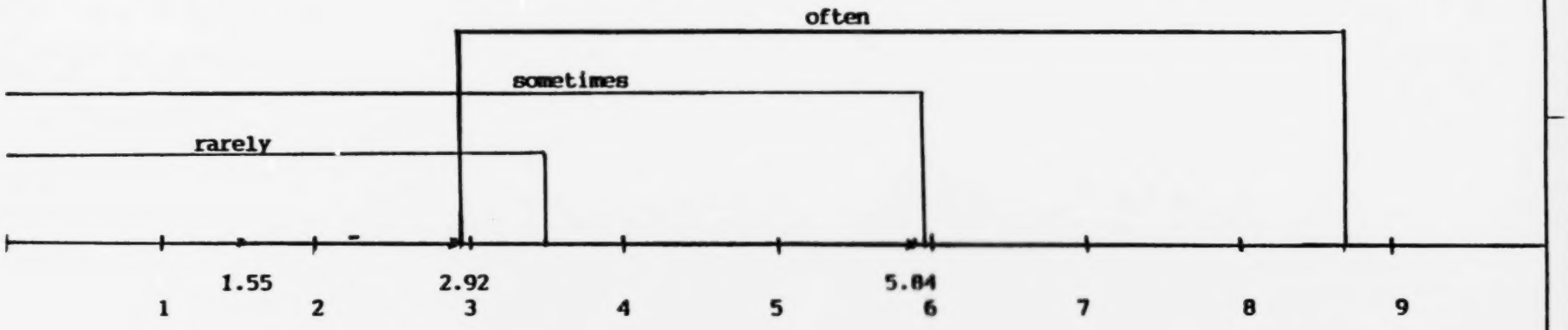
Table 10

Percent Change in "doing other things" by "full, some, or not much" Labels
1987 Sample

	Mean	Difference	Percent Change
Full	1.64		
		0.72	31%
Some	2.36		
		0.47	17%
Not Much	2.83		
Total	2.03	0.60	24%

FIGURE 1

Numerical Estimations of Times per Week Discussing Politics by Often, Sometimes, Rarely Labels:
Means and Range Defined by ± One Standard Deviation (1987 Sample)



Numerical estimations of Times per Week discussing Politics by Often, sometimes,rarely labels:
Means and Range Defined by ± One standard deviation (1984 sample)

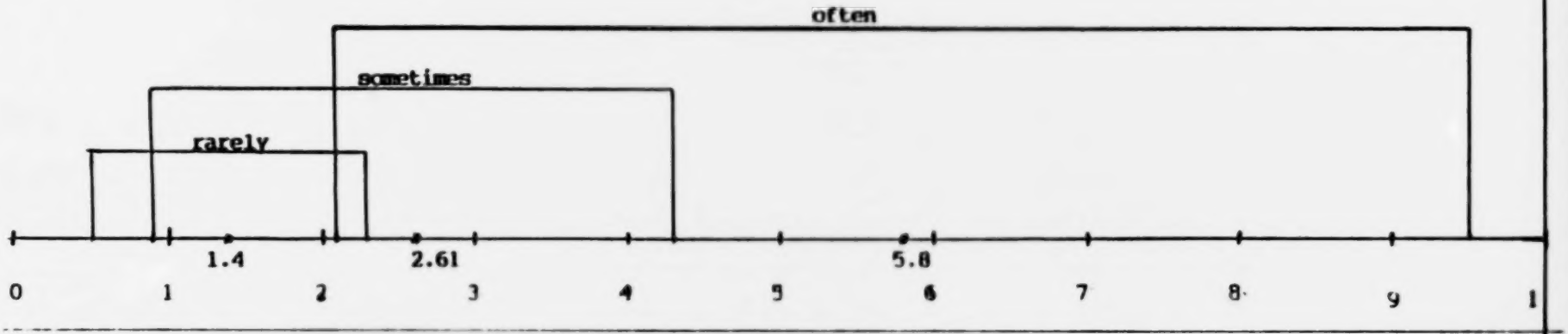


FIGURE 2

Discussion Index by Discussing Politics Often Sometimes, Rarely, Never Labels:
Means and Ranges Defined by ± One Standard Deviation (1984 Sample)

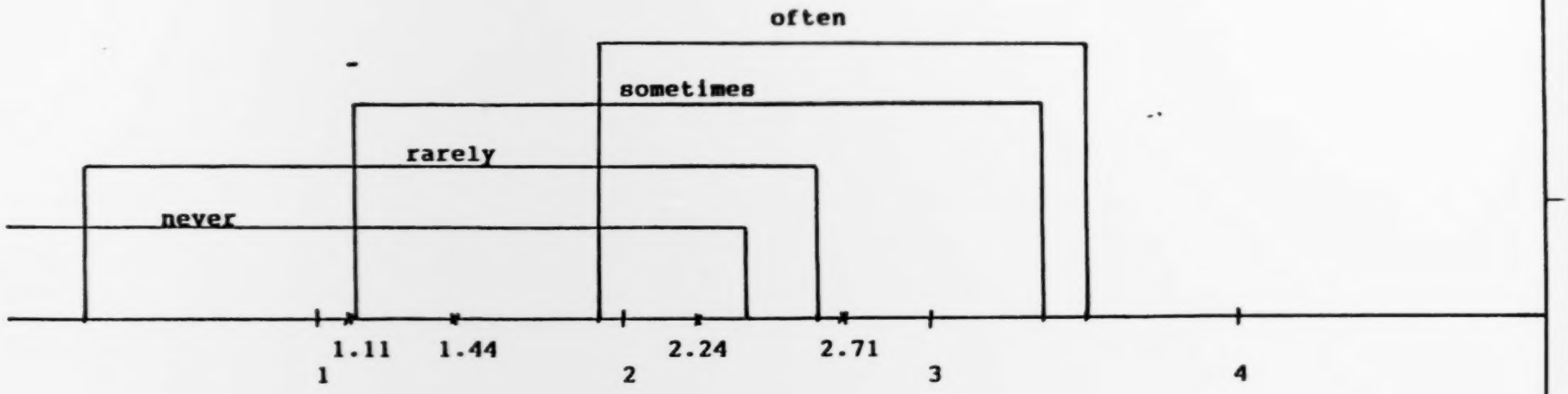
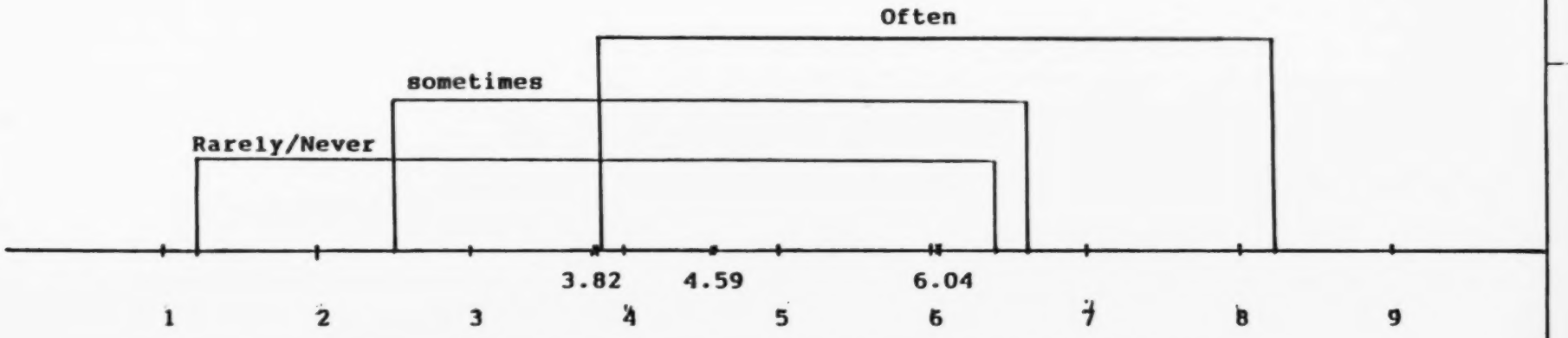


FIGURE 3

Amount of Front Page Read (1 - 10) by Reading About National and World Events Often, Sometimes, Rarely/Never: Means and Ranges Defined by ± One Standard Deviation (1987 Sample)



Days in Past Week Read Front Page by Reading about National and World Events Often, Sometimes, Rarely/Never: Means and Ranges Defined by ± One Standard Deviation (1987 Sample)

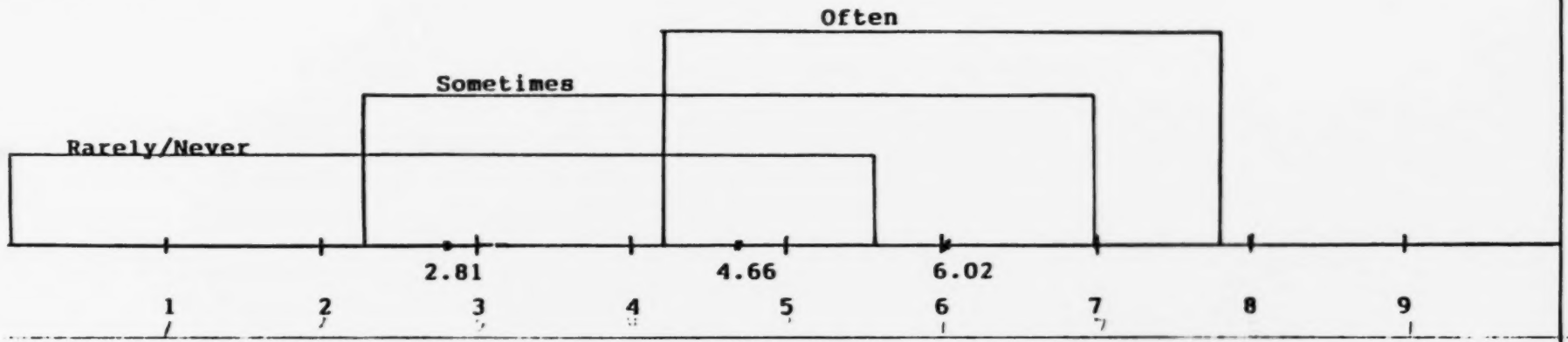


FIGURE 4

Frequency of "Doing Other Things" by Full, Some or Not Much Attention Paid to Television News: Means and Ranges Defined by + One Standard Deviation (1987 Sample)